

Evaluating Automatic Approaches for Word Meaning Discovery and Disambiguation using Lexical Substitution

Diana F. McCarthy
University of Sussex

Abstract

There has been a surge of interest in Computational Linguistics in word sense disambiguation (WSD). A major catalyst has been the SENSEVAL evaluation exercises which have provided standard datasets for the field over the past decade. Whilst researchers believe that WSD will ultimately prove useful for applications which need some degree of semantic interpretation, the jury is still out on this point. One significant problem is that there is no clear choice of inventory for any given task, other than the use of a parallel corpus for a specific language pair for a machine translation application. Most of the datasets produced, certainly in English, have used WordNet. Whilst WordNet is a wonderful resource it would be beneficial if systems using other inventories could enter the WSD arena without the need for mappings between the inventories which may mask results. As well as the work in disambiguation, there is a growing interest in automatic acquisition of inventories of word meaning. It would be useful to investigate the merits of predefined inventories themselves, aside from their use for disambiguation, and compare automatic methods of acquiring inventories. In this talk I will discuss these issues and some results in the context of the English Lexical Substitution Task, organised by myself and Roberto Navigli (University of Rome, “La Sapienza”) earlier this year under the auspices of SEMEVAL.